

Applying the Optimal Estimation Method for Retrieving Rayleigh-Scatter Lidar Temperatures in the Mesosphere

Jonathan Price¹, Vincent Wickwar¹, Robert Sica², Ali Jalali²

¹Utah State University, Logan, UT, USA

²University of Western Ontario, London, Ontario, Canada

Abstract

The Rayleigh-scatter lidar (RSL) system at the Atmospheric Lidar Observatory at Utah State University (ALO-USU) provided a rich database of absolute temperatures throughout the mesosphere from 45 km to above 90 km between 1993 and 2004. Recently, a new method for retrieving absolute temperatures from RSL observations has been developed by a group at the University of Western Ontario (UWO), Canada. The Optimal Estimation Method (OEM) uses machine learning to minimize a cost function by optimizing the temperature parameter in a forward model, in our case the lidar equation, to RSL data. This optimization provides some benefits over the existing method through a robust uncertainty budget and a quantitative determination of the cut-off altitude, or the topmost altitude in the temperature profile. Using this method also provides a slight increase in the top observable altitude and does not have a large dependence on the initial temperature. The OEM procedure was converted from MATLAB, which is used by the UWO group, into Python, which is used at ALO-USU. The temperatures were then reduced using the OEM from observations made between 1993 and 2004. Initial results obtained using the Python version of OEM were compared with those using MATLAB showing good agreement. More observations from ALO-USU were then

reduced using OEM and compared with the original reduction method. The results show good agreement between the two methods until higher altitudes. These differences can be attributed to dependence on initial conditions in the original method or over-constraining from overestimating the altitude range to be used in the OEM retrieval. At higher altitudes, however, the temperatures tend to agree within the given uncertainties. Further work with this method is being done to generate a temperature climatology using ALO-USU observations and developing a method to retrieve absolute neutral densities using a modification of the forward model in the OEM.

Introduction

Rayleigh-scatter lidar (RSL) is an important technique for obtaining temperature measurements throughout the middle atmosphere. This is because the region of atmosphere from the upper stratosphere (40 km) to the lower thermosphere (120 km) is difficult to observe. There are few instruments capable of observations in this region and fewer still that can observe the entire altitude range. The RSL at Utah State University is one such instrument capable of observing the range of the middle atmosphere. The original lidar operated at USU from 1993 until 2004, resulting in over 900 nights of observations, covering from 45

km to above 90 km. An upgraded system was made operational in 2012 which increased the observation altitudes to above 115 km. The current lidar system is one of the most powerful in the world. It uses two high power Nd:YAG lasers with the receiver consisting of four 1.25 m mirrors and the repurposed 40 cm mirror from the original lidar. The extended range of the new system covers from 40 km to above 115 km.

RSL temperatures are typically retrieved using the method outlined by Hauchecorne and Chanin (HC) (1980). Recently, a new method of temperature retrieval was developed for use with RSL observations using an Optimal Estimation Method (OEM) by Sica and Haefele (2015). This method improves on the HC method by providing a complete uncertainty budget, a mathematically represented cutoff altitude and can extend observations slightly higher in altitude all with less dependence on the a priori temperature profile.

The code developed to apply the OEM was developed in MATLAB. The processing has now been translated into Python at USU. To test the output of the Python code, a night from USU observations was processed using both programs to ensure consistency.

Theory

The OEM, described in Rogers (2011), uses machine learning to optimize the fit of a forward model to observed data by minimizing a cost function based on the parameter studied, in our case the temperature. The forward model used is the lidar equation, given by:

$$N(z) = \psi(z) \frac{n(z)}{z^2} + B, \quad (1)$$

where N is the number of observed photocounts, n is the absolute number

density, z is the altitude and B is the background noise. The instrument function, ψ , depends on factors such as detector area, atmospheric transmission, detector efficiency, and number of transmitted photons. To retrieve temperature, we solve for n using a combination of the ideal gas law and hydrostatic equilibrium. This gives:

$$n(z) = \frac{n_{top} T_{top}}{T(z)} * \exp\left(-\frac{1}{R} \int_z^{z_{top}} \frac{M(z')g(z')}{T(z')} dz'\right), \quad (2)$$

where T is the temperature, R is the gas constant, M is the mean molecular mass, g is the acceleration due to gravity.

The general form of the forward model, F , is:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \epsilon, \quad (3)$$

where \mathbf{y} is the measurement vector, the forward model depends on the state vector, \mathbf{x} , and the model parameters, \mathbf{b} , and ϵ is the measurement noise.

The optimal estimation for the retrieved state is obtained by minimizing a cost function. The cost function is formed with the measurement, \mathbf{y} , and its covariance, \mathbf{S}_y , the forward model, the retrieved state model parameters with the a priori covariance, \mathbf{S}_a (Sica and Haefele, 2015). The function is:

$$cost = [\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})]^T \mathbf{S}_y^{-1} [\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})] + [\hat{\mathbf{x}} - \mathbf{x}_a]^T \mathbf{S}_a^{-1} [\hat{\mathbf{x}} - \mathbf{x}_a]. \quad (4)$$

The a priori, \mathbf{x}_a , is obtained from the Naval Research Lab's (NRL) MSISe00 empirical model while the retrieved state, $\hat{\mathbf{x}}$, is obtained using the Marquardt–Levenberg method. As the cost function is minimized,

approaching 1, the fit is maximized resulting in an optimized temperature profile.

Along with the resulting temperatures, we derive an averaging kernel matrix, \mathbf{A} . This matrix is a useful diagnostics tool for determining how the retrieval reacts to a change in the real atmosphere. A perfect retrieval means the resulting temperature perfectly reflects the real atmosphere. This would show in \mathbf{A} being equal to the identity matrix (Rogers, 2011). In practice, the contribution from the a priori temperature will increase with altitude. By taking the trace of \mathbf{A} we can see how many degrees of freedom are available in the resulting temperature profile. In doing this, we obtain a value for the cutoff altitude as the bin number in the altitude array associated with the number of degrees of freedom. Above this altitude the a priori plays a significant role in the temperature retrieval.

Results

To make sure the translation of the OEM process from MATLAB into Python was done correctly, temperature results from both were plotted and compared. Temperatures from USU observations using the MATLAB version of the process (Figure 1a) were retrieved at UWO.

The same night was processed at USU using the Python version and the resulting temperature profiles are nearly identical (Figure 1b), except at higher altitudes where the a priori temperatures become more significant. The OEM temperatures also reproduce the HC results nicely, especially at altitudes below 85 km. The consistency between the MATLAB and Python results offer confidence in the quality of the Python version of the OEM. Because of this, more nights from USU were processed using OEM and compared with the original HC results.

Figure 2 shows the temperature results from OEM and HC for the night of 11/1/1996. Good agreement between methods can be seen with only small differences between the two.

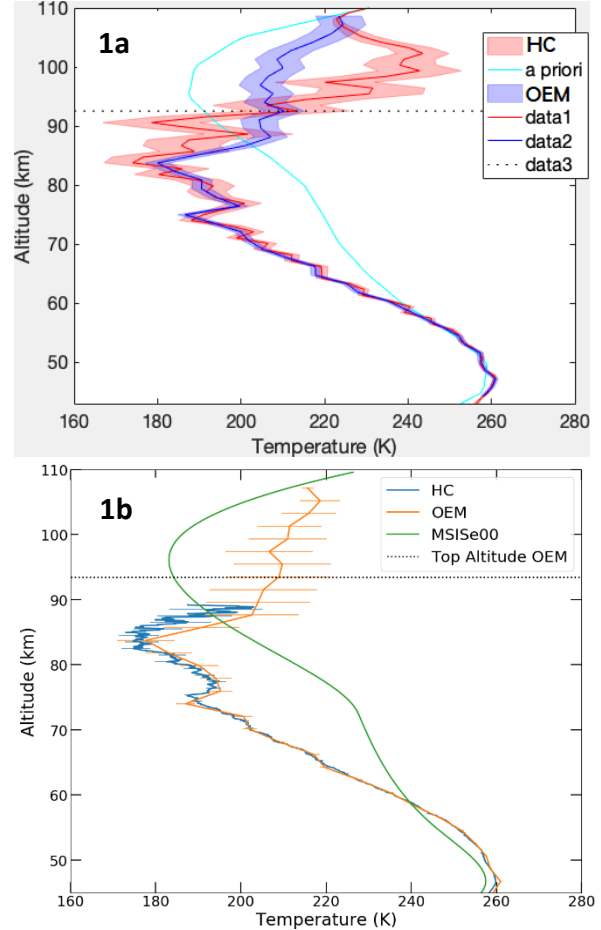


Figure 1a: UWO MATLAB temperature reduction of USU RSL observation from 9/14/2001. The blue line shows the OEM temperature, the red line shows the HC temperature. The cyan line shows the a priori temperature used in the OEM, UWO uses temperatures from the US Standard Atmosphere while USU uses MSISe00 (bottom, green). Figure 1b: The USU Python temperature reduction of the same night. The OEM temperatures (orange) are nearly identical to those with MATLAB (top, blue). Differences above 80 km are due to different a priori temperatures.

The total uncertainty budget for this night is shown in Figure 3. The dominant portion of the uncertainty comes from the statistical uncertainty related to the optimal fitting method. The calibration constant, which estimates the instrument parameters such

as efficiencies and photon count, becomes important around 90 km.

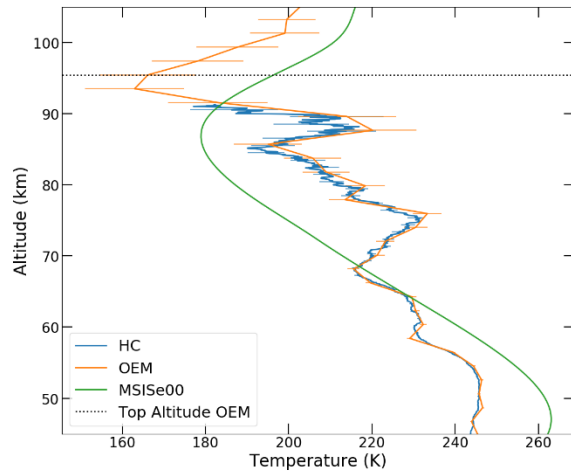


Figure 2: RSL USU temperature for 11/1/1996. HC and OEM show good agreement in temperature results with OEM gaining almost 5 km over HC.

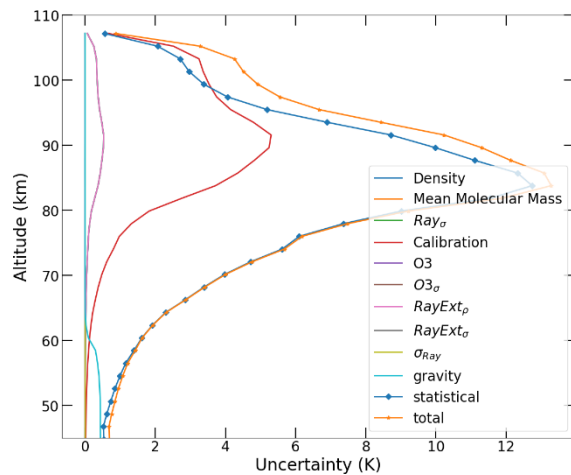


Figure 3: Uncertainties from OEM temperature retrieval for 11/1/1996. The dominating contributor is the statistical uncertainty with the calibration constant becoming somewhat significant around 90 km.

Some nights have more substantial differences in temperature results, such as on 4/4/1995 (Figure 4). Here we see temperatures diverging, with the HC lower than the OEM, above 70 km. Figure 5 is a plot of the residuals, the difference between the model output and the original photon counts. The residuals show no bias until around the cutoff altitude and increase in a

normal manner according to Poisson statistics suggesting a good temperature return. While it is not yet clear why such large differences are being produced between HC and OEM, this is being investigated.

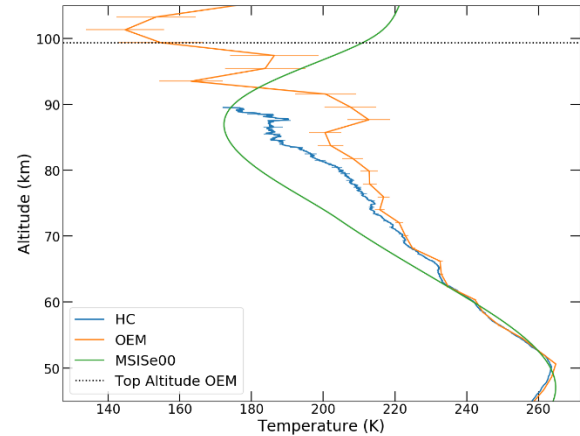


Figure 4: Temperatures from 04/04/1995. The HC temperatures diverge from the OEM starting above 70 km.

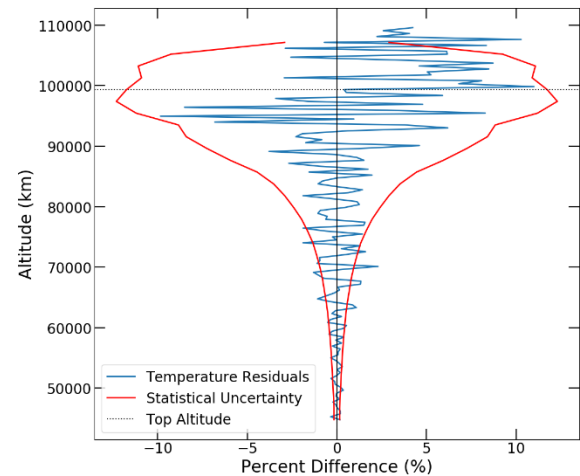


Figure 5: Residuals of OEM temperature reduction. This shows the percent difference between the model fit of the photocounts and the actual observed photocounts. The difference increases within the uncertainty according to Poisson statistics with no bias until around the cutoff altitude.

Discussion

Initial comparisons of retrieved temperatures between the original MATLAB

code and the new Python code show good agreement. This indicates the OEM temperature retrieval in Python is working correctly.

The differences between HC and OEM temperatures, in some cases, are quite different. Differences are expected but mostly at the highest altitudes rather than starting in the middle of the range. In the analysis by Jalali et al (2018), the authors show there are differences to be expected between methods but mostly at the topmost altitudes. This study, however, compares the temperature climatologies and not individual profiles. The extensive averaging used to produce the climatologies could cause the differences to appear smaller and so may not reflect the individual night differences accurately.

One potential cause for the differences can be due to the seeding temperature in the HC method. The analysis uses a seed temperature at the top altitude, often from a model such as MSIS00, in the temperature derivation. As the altitude decreases, the uncertainty for the temperature retrieval decreases as the dependence on this initial value decreases. This could be biasing the temperature to the initial value causing the larger differences at lower altitudes for some nights.

Another cause could be from overestimating the altitude range to be covered in the OEM analysis. The top altitude is arbitrarily selected at 110 km for the analysis. This value is chosen to ensure that the OEM will cover all portions of usable signal. However, over-constraining can affect the temperature profile at higher altitudes. To

test this, we can adjust the altitude range based on the cutoff altitude retrieved, such as making the range 5 km higher than the cutoff altitude. This could help with any over-constraining problems that might be present and will at least help us narrow down the causes for the larger differences between temperature reduction techniques. More comparisons with the UWO group are planned to help with this problem.

Conclusions

An optimal estimation method for retrieving Rayleigh-scatter lidar temperatures was developed in Python at USU. This method provides a full uncertainty budget and allows for a quantitative determination of the topmost obtainable temperature. Initial comparisons with results from the MATLAB code applied by the group at the University of Western Ontario show good consistency and suggests the translation into Python is good. Differences in some nights between the original temperature reduction method, HC, and the OEM merit investigation. Once the reason for the larger discrepancies in temperature are clear, work can be done in producing a temperature climatology of USU data. This can then be compared with the climatological study done by Herron (2007).

Future work is planned to apply the OEM to retrieve the absolute neutral densities in the mesosphere. This will be done by modifying the forward model to be optimized using number density rather than temperature.

References

Hauchecorne, A., Chanin, M., (1980), Density and temperature profiles obtained by lidar between 35 and 70 km, *Geophysical Research Letters*, 7(8), 565-568.

<https://doi.org/10.1029/GL007i008p00565>

Herron, J. P., (2007), Rayleigh-Scatter Lidar Observations at USU's Atmospheric Lidar Observatory (Logan, UT) – Temperature Climatology, Temperature Comparisons with MSIS, and Noctilucent Clouds, *All Graduate Theses and Dissertations*. 4686

<https://digitalcommons.usu.edu/etd/4686>

Jalali, A., Sica, R. J., and Haeferle, A. (2018), Improvements to a long-term Rayleigh-scatter lidar temperature climatology by using an optimal estimation method, *Atmospheric Measurement Techniques*, 11(11), 6043-6058.

<https://doi.org/10.5194/amt-11-6043-2018>

Rogers, C. D. (2011), Inverse Methods for Atmospheric Sounding: Theory and Practice, *World Scientific Publishing Co.* ISBN: 9789812813718.

<https://doi.org/10.1142/3171>

Sica, R. J., Haeferle, A. (2015), Retrieval of temperature from a multiple-channel Rayleigh-scatter lidar using an optimal estimation method, *Applied Optics*, 54(8), 1872-1889.

<https://doi.org/10.1364/AO.54.001872>

Wickwar, V. B., et al (2001), Mesospheric Temperature Observations at the USU/CASS Atmospheric Lidar Observatory (ALO), Remote Sensing of the Atmosphere, Environment, and Space, 9-12 October 2000 Sendai, Japan.